



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Loss given default models incorporating macroeconomic variables for credit cards

Citation for published version:

Crook, J & Bellotti, T 2012, 'Loss given default models incorporating macroeconomic variables for credit cards', *International Journal of Forecasting*, vol. 28, no. 1, pp. 171-182.
<https://doi.org/10.1016/j.ijforecast.2010.08.005>

Digital Object Identifier (DOI):

[10.1016/j.ijforecast.2010.08.005](https://doi.org/10.1016/j.ijforecast.2010.08.005)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

International Journal of Forecasting

Publisher Rights Statement:

This is an Author's Accepted Manuscript of the following article: © Crook, J. & Bellotti, T. Jan 2012, "Loss given default models incorporating macroeconomic variables for credit cards", in *International Journal of Forecasting*. 28, 1, p. 171-182. The final publication is available at <http://dx.doi.org/10.1016/j.ijforecast.2010.08.005>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Loss Given Default Models incorporating Macroeconomic Variables for Credit Cards

1. Introduction

Loss Given Default (LGD) is the loss incurred by a financial institution when an obligor defaults on a loan, given as the fraction of exposure at default (EAD) unpaid after some period of time. It is usual for LGD to have a value between 0 and 1 where 0 means the balance is fully recovered and 1 means total loss of EAD. LGD is an important value that banks need to estimate accurately for several reasons. Firstly, it can be used along with probability of default (PD) and EAD to estimate expected financial loss. Secondly, a forecast of LGD for an individual can help determine the collection policy to be used for that individual following default. For example, if high LGD is expected, then more effort may be employed to help reduce this loss. Thirdly, an estimate of LGD, and therefore portfolio financial risk, is an integral part of the operational calculation of capital requirements to cover credit loss during extreme economic conditions. The Basel II Capital Accord [2006] allows banks the opportunity to estimate LGD using their own models with the advanced internal ratings based (IRB) approach.

In this paper we focus on modelling and forecasting LGD for UK retail credit cards based on account variables (AVs) and also the inclusion of macroeconomic variables (MVs). Our prior expectation is that as interest rates rise, so the cost of mortgage and other debt increases, making it more difficult for an obligor to repay outstanding

credit card balances and so increasing the mean LGD. Equally, an increase in unemployment level means more people find themselves in circumstances where they cannot repay credit and this would also increase the mean LGD. On the other hand, an increase in earnings would mean more people have more income available to pay off debt and would therefore decrease mean LGD. In addition it is possible that some defaulters are more likely to be less able to repay than others when the state of the economy changes. For example, those who are unemployed at the time of credit card application may be particularly sensitive to interest rate increases, as may home owners. Similarly borrowers with a higher default balance may be particularly sensitive to increases in interest rates. For this reason we also consider interactions between MVs and account data. We consider four key research questions:-

Q1.Which credit card application and default variables are the key drivers of retail LGD?

Q2.What is the best modelling approach for retail LGD?

Q3.How well do the models perform at forecasting LGD?

Q4.Does the inclusion of MVs lead to improved models of retail LGD?

We investigate these questions by building several alternative models of LGD. We find that there are many important drivers of LGD taken from application details and default information. Given that the distribution of LGD is a bimodal U-shape, we consider a Tobit model and a decision tree model along with various transformations of the dependent variable. Although LGD is not easy to model and poor model fit is typical, nevertheless we find that models can be built which provide improved estimates of LGD and good forecasts of mean LGD across a portfolio of accounts.

Surprisingly, we find that the best forecasting model is Ordinary Least Squares (OLS) regression. Economic conditions are included as values of MVs for bank interest rate, unemployment level and earnings growth at time of account default. We find that the first two MVs are statistically significant explanatory variables that give rise to improved forecasts of LGD in hold-out tests at both an account and portfolio level. Building LGD models with MVs also addresses the Basel II requirement to estimate “downturn LGD” since stressed values of MVs can be used in the model to forecast LGD during poor economic conditions. This can be done by stressing interest rate values as we explain in our conclusions.

The modelling and forecast of LGD for retail credit using macroeconomic conditions is a new area of study. There is an extensive literature regarding LGD models for corporate loans (see eg Altman et al [2005]). However, there is less about *forecasting* LGD. An exception is Gupton and Stein [2005] who describe a predictive LGD model for corporate loans using Moody-KMV’s Losscalc[®] software. There is also very little literature regarding retail credit LGD, even though this is a large financial market: total lending in the UK consumer credit market reached over £1.4 trillion in 2009 [source: Bank of England]. Grippa et al [2005] publish empirical LGD models for a sample of 20,724 Italian accounts that includes small businesses along with households. They observe differences in LGD and recovery periods across different geographic regions and different recovery channels. They also conducted a multivariate analysis that showed a statistically significant negative relationship between the presence of collateral or personal guarantee and LGD, and a positive relationship with size of loan. However, the range of variables used is far more limited than would be available to a financial institution that has made credit card or

personal loans and the study did not attempt to forecast LGD. Dermine and de Carvalho [2005] model LGD for loans to small and medium sized firms in Portugal. They apply mortality analysis and include annual GDP growth as an explanatory variable. However, they found that GDP growth was not significant. They suggest that this may be due to the fact that the period of analysis, 1995 – 2005, did not include a significant recession. We may also note that their training sample size (374 defaults) was relatively small and may not have been large enough for a significant relationship between the economy and LGD to be discovered. Querci [2005] provides an LGD model for loans to small businesses and individuals by an Italian bank. This study shows the importance of regional differences on LGD variation but does not include time varying macroeconomic conditions. Figlewski et al [2007] model the effect of macroeconomic factors on corporate default with a detailed study of numerous economic conditions including unemployment level, inflation, GDP and a production index. They found that many of these MVs were significant explanatory variables. Saurina and Trucharte [2007] model PD for retail mortgage portfolios in Spain. They show that the GDP growth rate is a significant cyclical variable in the regression and has a negative sign as we would expect. That is, during downturns (low GDP growth), PD increases. However, they also include an interest rate variable and, although it has a positive sign and is significant, they report that including interest rates does not improve accuracy.

The novelties of our paper are that unlike published work (1) we consider forecasts of LGD for retail credit cards, (2) we report results of model comparison, (3) we include macroeconomic conditions in our models and (4) we do this using a very large sample across several different credit card products. In section 2 we describe our

modelling and performance assessment methods. In section 3 we discuss the application and macroeconomic data used. In section 4 we provide model comparisons and test results along with a description of an explanatory model with MVs. Finally in Section 5 we provide some conclusions and discussion.

2. Method

We consider several models as combinations of different variables, modelling frameworks and data transformations.

2.1 Models

In general, for retail credit, there are five categories of circumstances that will affect the amount an individual repays on a defaulted loan and can be used to build models of LGD:

- (1) individual details, some of which can be collected at time of application such as age, income, employment, housing status and address;
- (2) account information at default: date or age of account at default and outstanding balance;
- (3) changes in personal circumstances of an obligor over time;
- (4) macroeconomic or business conditions on date of default, or possibly with a lag or lead on date of default;
- (5) operational decisions made by the bank, such as the level of risk they were willing to accept on the credit product and the process they use to follow up bad debt.

Of these, the richest source of explanatory variables we have is the information provided at time of application for credit along with the credit bureau score collected

by the bank at time of application. This is data falling into category (1). We also include category (2) data, account information at default. Including this data implies the model is conditional on default. It is possible to build models unconditionally but this is outside the scope of this paper. It is difficult for a lender to extract data in category (3). It cannot easily keep track of an individual's employment status or, even less so, his or her personal difficulties, such as divorce or illness, that may lead him or her to be unable to fully repay debt. It is possible to use account behaviour data or a behavioural score but we do not do this in this study since such information is not homogeneous within the data we have. It is understood that LGD is likely to be time dependent, varying over the business cycle [Schuermann 2005]. Therefore we include macroeconomic conditions (4). Including a bank's operational decisions (5) for each credit card product over time could also be fruitful. However, this information was not available for our study.

To answer Q4, we build and compare models with and without MVs. The question can be further explored by contrasting with models that also include interaction terms between AVs and MVs. To help answer Q3 we should contrast our models with a simple model built with no variables. Within the context of OLS this simple model effectively forecasts LGD to be the mean value taken from the training data set. If the more complex models give improved forecasts over this simple model then the variables we include provide useful information for LGD estimation. Therefore we have four model structures based on including different explanatory variables:-

- **Simple:** no covariates in the model.
- **AV:** account variables only.
- **AV&MV:** account and macroeconomic variables.

- **AV&MV with interactions:** also includes interaction terms between AVs and MVs. It is not feasible to include all interaction terms so variable selection is used as described later in Section 3.3.

We do not restrict our models to OLS but also consider three alternatives. Tobit and a decision tree model are considered since they have a structure better suited to the bimodal nature of LGD. Least absolute value regression is also considered since it may be that the absolute error may be a more sensible criterion for estimating LGD than least square error.

Since LGD follows a truncated distribution with a large number of cases at the extreme values 0 and 1, the Tobit model may be better suited since it takes account of bounds on a dependent variable y_i through truncation. The two-tailed Tobit model uses a latent variable y_i^* to model boundary cases such that $y_i^* = \beta \cdot \mathbf{x}_i + \varepsilon_i$ where $y_i = \min(1, \max(0, y_i^*))$. Assuming the distribution of the residuals conditional on \mathbf{x} is normal, the following log-likelihood function is constructed for maximum likelihood estimation of β and variance of residuals σ^2 :

$$\log(\beta, \sigma) = \sum_{0 < y_i < 1} \log \left[\phi \left(\frac{y_i - \mathbf{x}_i^T \beta}{\sigma} \right) / \sigma \right] + \sum_{y_i = 0} \log \left[1 - \Phi \left(\frac{\mathbf{x}_i^T \beta}{\sigma} \right) \right] + \sum_{y_i = 1} \log \left[\Phi \left(\frac{\mathbf{x}_i^T \beta - 1}{\sigma} \right) \right]$$

where ϕ and Φ are the probability and cumulative density functions for the standard normal distribution respectively. This likelihood function is constructed by considering the probabilities of the dependent variable being between the boundaries and also on each boundary separately [Greene 1997, pp. 962-966].

The decision tree model uses two logistic regression sub-models to model the special cases for total loss and no loss, ie $LGD = 1$ and 0 respectively, as binary classification problems. Then, if $0 < LGD < 1$, an OLS regression model is used. This decision tree model is illustrated in Figure 1. We may expect this decision tree to be good at modelling LGD due to the large number of boundary cases at 0 and 1 which allow us to naturally approach the problem as a hybrid of two classification problems and a regression problem. This approach is meaningful since we may suppose there are special conditions which would make a customer pay back the full amount of debt or to pay back nothing, rather than just a portion. LGD is forecast for an account i as the expected value given the three sub-models, ie $(1 - p_{0i})(p_{1i} + (1 - p_{1i})L_i)$ where p_{0i} is the probability of $LGD=0$ for account i computed from the second estimated logistic regression model, p_{1i} is the probability of $LGD=1$ from the first estimated logistic regression model and L_i is the estimate of LGD assuming the loss is fractional and computed from the regression model.

FIGURE 1 HERE

As is conventional in the literature we model LGD in terms of *recovery rate* (RR) rather than LGD directly, where $RR = 1 - LGD$. Our working definition of RR is

$$RR = \frac{\text{sum of repayments made over a period } t \text{ following default}}{\text{outstanding balance at date of default}}.$$

The choice of recovery period is a business decision. We consider a recovery period of $t=12$ months following default. Often banks are interested in longer periods or

want to estimate recovery at close of account, but as we discover, a 12 month model can be used to estimate for longer recovery periods. It is possible to calculate LGD in alternative ways. For example, time after default could be an event, such as account charge off, rather than a fixed period; or the market value of the bad debt based on sale of the exposure in the market could be taken into account along with repayments; or administration costs of following up default may be included in the LGD calculation. However, these alternatives are beyond the scope of this paper. Loss of interest payments could also be included in the definition of LGD but this is not required by Basel II.

Since the distribution of RR is bimodal and U-shaped, we also consider modelling fractional logit, beta distribution and probit transformations of RR.

Fractional logit transformation: $T_{RR} = \log(RR) - \log(1 - RR)$.

The fractional logit model is particularly attractive since it deals specifically with response variables in the range 0 to 1, like RR, by transforming them into a larger range of values. It has been applied in several other econometric analyses [Papke & Wooldridge 1996] and was applied in particular by Dermine and Neto de Carvalho [2005] for modelling LGD for corporate loans.

Beta distribution transform: $T_{RR} = \Phi^{-1}(\text{beta}(RR, \alpha, \beta, 0, 1))$ where Φ is the cumulative density function of the standard normal distribution and α, β are parameters estimated from training data using maximum likelihood estimation. The Beta distribution is appealing since it is able to model bimodal variables with a U-shaped distribution over the interval 0 to 1. It is therefore particularly useful for RR

and tends to transform RR into an approximately normal distribution. Figure 2 shows a distribution of the Beta transformation of RR for the credit card default data we use. It illustrates how the transformation creates an approximately normal distribution between two extreme values. The Beta distribution has been used successfully in Moody's KMV Losscalc[®] software package for modelling RR [Gupton & Stein 2005].

FIGURE 2 HERE

Probit transformation : $T_{RR} = \Phi^{-1}\left(\frac{|\{i : R_i \leq RR\}|}{n}\right)$ where Φ is the cumulative density function of the standard normal distribution and R_1, \dots, R_n are observed RRs taken from training data. This transformation uses a nonparametric approach to transform RR into a normal distribution based on the empirical distribution in the training data.

2.2 Model assessment

For OLS we report *adjusted* R^2 for model fit. This is for two reasons. Firstly, the various models we consider are nested so inevitably those with additional covariates will give improved R^2 model fit. The adjusted R^2 compensates for the additional variables. Secondly, reporting R^2 is misleading since it does not give a fair comparison between different studies with different sample sizes.

We report coefficient estimates for the OLS model with RR as dependent variable. However, since RR is not normally distributed, the error terms may not be normally distributed. Therefore conventional estimators of standard errors may not be

unbiased. Instead we use a bootstrap to construct distributions for the coefficient estimates [Kennedy 2003, section 4.6]. As Lam and Veall [2002] show when OLS is used with non-normal, and in particular bimodal, distributed error terms, the bootstrap gives accurate estimates of confidence intervals where the usual analytic method fails.

We report the relative effect of each MV within the model. The coefficient estimates for MVs are multiplied by their standard deviation over the training period to derive *standardized coefficient estimates*. They show change in RR for a one standard deviation change in the covariate value and are therefore comparable and give an indication of the relative importance of each MV within the model. This approach was taken to study the effects of MVs on corporate default by Figlewski et al [2007]. Since our credit card data spans the period from 1999 to 2005, standard deviations for each MV are computed based on values within this period.

2.2.1 Hold-out test procedure for forecasts

To test the effectiveness of the LGD model for forecasts, we use a hold-out sample, testing on credit card default data that is independent and follows chronologically after the period of the training data used to build the models. This approach allows us to simulate the expected operational use of LGD models in retail credit when a financial institution may want to assess the LGD risk on a new batch of defaults based on the performance of past defaults. In detail, we select cohorts of test data sets consisting only of accounts that default in a particular quarter. For each of these cohorts we train only on default data available prior to that quarter. Since we need to measure LGD after a period of t months following default, we need to ensure that date of defaults in the training data are at least t months prior to the beginning of the test quarter. This test procedure is illustrated in Figure 3. So, for example, if our test set

was 2003Q3 and we are considering LGD after 12 months, our training data would consist of all cases that default within the period from 1999Q1 to 2002Q2. To estimate robust models with MVs we should train over the whole business cycle which is usually considered between 3 to 5 years. Therefore we consider taking training data sets with a minimum of 3 years of defaults. This then gives us 10 quarters of test set data from 2003Q1 to 2005Q2. Results across these independent test sets then form a time series of forecast results.

FIGURE 3

The accuracy of forecasts relative to the observed true values is measured at the account level by mean square error (MSE). However, we are also interested in how well the model is able to estimate the observed, or true, mean LGD or RR over a portfolio of accounts. Therefore, for each test set quarter we measure the difference between forecast and observed mean RR across all test cases. If this difference is greater than zero, then the model is generally overestimating RR, whereas when it is less than zero, it is underestimating RR. The closer the difference between forecast and observed RR is to zero, the better the estimate. The mean value of both the MSE and absolute value of difference in forecast and observed mean RR (abs diff RR) across the several test quarters are reported in order to get an aggregate measure of performance.

Since RR must be between 0 and 1, we truncate all forecasts of RR to fall within that range prior to measuring performance for all cases and all models. Also, to generate comparable results, performance is measured between predicted and observed RR, regardless of which transformation of RR is modelled. Therefore, if a transformation

of RR is used, the inverse transform is applied to extract predicted RR and performance is measured on this rather than the transformed value. This is reasonable since ultimately the value we want to model is RR and the transform should merely be a means to that end.

2.2.2 Forecasting 24 month LGD using a 12 month LGD model

For these experiments, we will be assessing models of LGD after 12 months. However, financial institutions often follow a bad debt over several years, so they are also interested in forecasting LGD over a longer period of say 24 or 48 months. For this reason, we also assess the 12 month model to see how well it forecasts for longer periods in comparison with a 24 month model. If it does well, then this implies that a single LGD model may be used to forecast for any LGD period. A 12 month LGD model can be used to forecast 24 month LGD by calibrating the 12 month forecasts to 24 month forecasts. A simple way to do this is to use OLS regression on the 24 month LGD training data to build a linear model of 24 month RR with an intercept and 12 month RR as the explanatory variable. This model is then used to convert 12 month to 24 month forecasts. We use the same training and testing scheme as is set out above. For 24 month LGD we use 6 quarters of test data from 2003Q1 to 2004Q2.

3. Data

3.1 Application data

For this study we have available a data set consisting of over 55,000 credit card accounts in default over the period 1999 to 2005 for customers across the whole of the

UK. Account holders are expected to make a minimum payment of outstanding balance each month. We define *default* as a case where a credit card holder has failed to make minimum payments for three consecutive months or more. This is a typical definition of default for credit cards [Thomas et al 2002, p.123] and is in line with the default definition given by the Basel II Accord [BCBS 2006]. The data consists of four different credit card products which are a selection from those offered by a financial institution. As is typical for LGD, its distribution in our data set is between 0 and 1 and approximately U-shaped¹. The calculation of LGD should ideally include administration costs for managing and implementing a collection procedure following account default. Unfortunately, this information was not available for the credit card data we used. The credit card data we use has many details extracted at time of application. These include the applicant's housing and employment status, age, income, total number of known credit cards and length of relationship with bank (time with bank). A credit bureau score is also provided and is from the same source and so is homogeneous across products. Demographic information is provided to classify the areas of residence of the credit card holder at time of application. The demographic information was from the same source for all credit card products and was coded into four broad categories: (1) council or poor housing, (2) rural, (3) suburban or wealthy area and (4) others. Additionally, information at time of default is also included in the data. This consists of balance outstanding at default and age of credit card account. We include balance at default since there is strong evidence from past studies that it is an important effect [Grippa et al 2005; Dermine and de Carvalho 2006] and it makes sense to include it operationally, especially if it improves forecasts of LGD. Table 2 shows the full list of variables used. Some variables, time with bank, income and age, have a small percentage of missing values (less than 6% of accounts). For each of

these variables, we code missing values to 0 and create a dummy variable to indicate a missing value to capture the mean value amongst accounts with missing values.

Several studies have found that PD and LGD are positively correlated (see Rösche and Scheule (2006)). We also found this to be the case in our data. Nevertheless, PD is not included in our models since it is effectively represented by including the application variables that are usually used to model it, along with macroeconomic variables that may explain the joint systematic risk to both PD and LGD (Altman et al 2005).

Any single credit card portfolio is liable to have operational effects that will alter overall risk over time for that specific product, such as changes in cut-offs on credit score when accepting applications. This may lead to idiosyncratic links to economic conditions and therefore poorer models using MVs. By combining data across several products the impact of these idiosyncratic effects will be reduced and changes in risk over time are more likely to be linked to more objective effects such as the economy. Additionally, combining several products into one data set will increase the training set size. These two factors should lead to stronger MV models and we test this hypothesis by running experiments for all products combined and also for each product separately. When all products are included in the data set, a dummy variable is used to indicate which product the account belongs to, in order to model different levels of RR between products.

3.2 Macroeconomic variables

We consider these three series of macroeconomic data for the UK which we believe would have a strong direct effect on mean LGD for UK retail credit cards:-

- Selected UK retail banks' base interest rates.
- UK unemployment level: measured as thousands of adults (16+) unemployed .
- (Earnings) UK earnings index (2000 = 100) for the whole economy including bonuses as a ratio of the retail price index.

These are all available from the UK Office for National Statistics as monthly data. We use non-seasonally adjusted data for earnings since we expect that seasonal changes in the economy may have some effect on abilities to repay. We would also have preferred to use non-seasonally adjusted data for the unemployment level but unfortunately this was not available. GDP growth for the UK is a common indicator of economic conditions but we have not included it since it is not available as monthly data which is the granularity that lenders typically require and therefore the granularity we require for our models. The MVs are included for each case at the date of default but it may be that if there is a relationship between MVs and LGD then this is lagged or led. For example, changes in interest rates may, in general, affect ability to pay several months later. We experimented with several different lag lengths and found better performance for lags of 0 and 6 months. For this reason, we also consider LGD models with MVs lag or lead 6 months.

Each MV has a time trend: interest rates and unemployment level are generally falling over the period 1999-2005 whilst real earnings are steadily increasing. Indeed, earnings generally increase exponentially with time therefore we include it in our

model as growth in log earnings over 12 months to remove this obvious time trend. The three MV time series we use are shown in Figure 4. We ensure that any model fit of MVs as explanatory variables is not simply because they follow a time trend that matches a trend in RR by including date of default explicitly in the models. If a MV is a good explanatory variable simply because of a time trend, then the inclusion of date of default should weaken its effect within the model. Including date of default in the AV model also allows us to test whether improvements in forecasts are simply due to a time trend rather than specifically economic conditions.

FIGURE 4

Different effects on LGD over time could be captured by using dummy variables for cohorts either at the yearly or quarterly level. There are two reasons we do not do this however. Firstly, the freedom gained when using any time dummies could simply absorb the effect we expect the MVs to explain. Secondly, although fine for explanatory models, it is not clear how such time dummies could be used for forecasting on a hold-out sample since the dummy variables for the period of the hold-out cohort will necessarily have the value 0 for all cases in the training data and so will not have a coefficient estimate.

High correlation between MVs is a potential problem since this could lead to multicollinearity within the LGD model and therefore distort parameter estimates. We can test for multicollinearity by measuring the variance inflation factor (VIF) given by $(1 - R^2)^{-1}$ when each MV is regressed on all other model covariates (Kennedy 2003). A high VIF indicates multicollinearity and a VIF greater than 5 is an indication that there may be a problem.

3.3 Inclusion of interaction terms

Since there are many possible combinations of variables to form interaction terms, the number included in the model is controlled using forward variable selection. All AVs are included in the model but an iterative process is used to include MVs and interaction terms between AVs and MVs. At each step, each of the outstanding MVs and interaction terms not already in the model are added separately. The term that maximally increases a fit criterion is added to the model. The process is repeated until no new interaction terms are found that improve fit. There are several possible fit criteria that could be used and it is common to use an F-test. However, since we are interested in forecasting, we use Akaike's information criterion (AIC) [Akaike 1973]. This has the advantage that it takes account of the parameter space of the model and discourages complex models with large numbers of variables. In turn, this discourages over-fitting to the training data set. We approximate AIC by $n \ln(\text{MSE}) + 2p$ where n and p are number of observations and number of parameters in the model respectively and MSE is the mean square error for observations in the training data. We find using the AIC criterion gives better forecasts than using the standard F-test. Further discussion of variable selection methods and use of AIC for predictive models is given by Miller [1990]. The variable selection procedure we use is further constrained so that for each interaction term included, its constitutive terms are also automatically included [Brambor 2005].

4 Results

Section 4.1 describes forecast performance for comparison of the different models.

Then section 4.2 describes the best performing model for forecasts and its statistically

significant explanatory variables. Models for longer recovery periods are considered in the third section.

4.1 Model comparisons

Table 1 shows forecast results for different models. Focussing on the first section showing results for a 12 month recovery period and for all products combined, it is clear that the standard OLS model with both AVs & MVs performs best for both measures of forecast performance. The more complex models, Tobit, decision tree or least absolute value regression give worse performance and so does using any of the transformations of RR. We may expect OLS to do well for the MSE measure, but we may expect that one of the alternative models would be better when estimating mean RR over the portfolio. In particular we may expect least absolute value regression to be better since it is a linear loss function. However OLS forecasts best estimates of mean RR. This is a robust result which was obtained with many alternative experiments. The inclusion of interaction terms gives slightly worse forecasts than the AV&MV model so including interaction terms does not provide any benefit in estimating LGD. It is also notable that the simple model that effectively forecasts the mean RR from the training data set does well and outperforms many of the more complex models. Nevertheless using AVs along with MVs shows considerable gain in performance over the simple model.

TABLE 1

We also consider MVs with lags or leads of 6 months. We consider lags since it is possible that the effect of the economy on obligor behaviour may be delayed. We

consider a lead since this would give the values of MVs midway through the recovery period. Table 1 shows that using lags or leads on MVs gives worse forecasts than using MV values at time of default and are almost as bad as not including MVs at all.

The second section of Table 1 shows results for separate products. It is clear that the AV&MV model does not do consistently well for separate products and rarely as well as when all products are combined. In many cases the simple model with no explanatory variables is the best performing for forecasting mean RR. This is due partly to operational peculiarities within each portfolio that may have a spurious link to macroeconomic movements and also partly to reduced training sample size. These results suggest that better LGD models can be built when data from different products are combined.

Figures 5 and 6 show forecast time series results for the first three models listed in Table 1 in more detail for each test quarter. They show that the AV&MV model performs consistently better over time. Figure 5 shows that MSE is lower for AV&MV than either the simple or AV model and improves over time, possibly a result of having training data over a longer period of the business cycle to model the MV effects. Figure 6 shows that overall the AV&MV model forecasts mean RR much more closely than the other models as is evidenced by how close the difference between forecast and observed mean RR is to zero. In contrast the AV model is consistently over-estimating RR over time. In 2004Q2, the simple model achieves the best result but this is serendipitous since its forecasts are simply moving from underestimating to overestimating RR at that time. These results show that including MVs is important to improve LGD forecast results. It should be noted, however, that

in our study at least 3 years of training data is required. We found that using less than this led to unstable models with MVs that occasionally extremely over or under estimate LGD.

FIGURE 5

FIGURE 6

4.2 Explanatory model

Table 1 shows that the AV&MV model using OLS regression was the best performing forecaster so we describe this model estimate in further detail. We report model fit results for LGD models built on all data from 1999 to 2005. Including MVs in the LGD model improves fit to training data and is statistically significant. The AV model has an adjusted R^2 model fit of 0.105. When MVs are included this increases to 0.110. When interaction terms between MVs and application variables are also included using forward selection, the adjusted R^2 is 0.111. This small increase indicates that adding interactions does not give a noticeable improvement reinforcing the results observed for forecasting. Table 2 shows coefficient estimates using OLS regression with the AV&MV model. Since the error residuals are non-normal we have used bootstrap to compute statistical significance. The reported p-values are from a normal-based distribution imposed on bootstrap coefficient estimates. This is reasonable since we found the 95% confidence intervals for the normal-based distributions matched closely those for the empirical percentile distribution sharing never less than 92% of the others' range.

TABLE 2

Many of the model variables are statistically significant at a 0.01 level. In particular, housing status is important with council and private tenants having generally lower RR than home owners; the longer the time that the customer has been with the bank (time with bank) and the longer the individual has held the credit card prior to default (time on books at default) both tend to increase RR; individuals with higher income also tend to have higher RR. All these are indicators of *customer stability* which we would expect to give lower risk. Higher credit bureau scores tend to give higher RR, which again shows that individuals with expected low credit risk tend to pay back more of their bad debt. Also, the size of the initial balance at default has a negative effect on RR which is what we would expect since higher outstanding debt is more difficult to pay back. We note a positive correlation between default balance and income, which suggests that people with high income tend to build up larger balances on their credit card. Indeed if we remove default balance from the model the sign on income becomes negative since it becomes a surrogate for the missing default balance variable. However, when both income and default balance are in the model together then the sign on income becomes positive which is what we would expect since the availability of higher income implies a greater capacity to repay the outstanding balance. Employment status has less impact, although we see that home makers tend to have lower RR. Demographic information was important with those living in areas classified as council or poor housing tending to have lower RR than those in rural, suburban or wealthy areas.

Table 2 shows that coefficient estimates for MVs have the expected signs. That is, the parameter estimate for interest rates is negative meaning that higher interest rates

at the time of default tend to give lower RR. Similarly, higher unemployment levels are also linked to lower RR. However, higher earnings growth, year-on-year, leads to increased RR which suggests earnings growth leads to better recoveries. Bank interest rates and unemployment level are both statistically significant at a 0.01 level, although earnings growth is not statistically significant in the model. The coefficient estimate for date of default is statistically significant but has a relatively small effect when its annual effect is compared with standardized estimates for the significant MV coefficients as shown in Table 2. This implies that the effect of the MVs is not due to a simple time trend. Bank interest rate clearly has the largest magnitude with unemployment level having less than half the interest rate effect. Additionally we find that the VIF for any of the MVs when regressed on all other covariates in model (1) was always less than 2 which is sufficiently small that we should not expect that the results are affected by multicollinearity. Since including interaction terms did not improve forecasts or model fit we do not report interaction terms in the explanatory model.

4.3 Forecasting 24 month LGD with a 12 month LGD model

We test whether it is possible to use a model built for a 12 month recovery period to forecast for a 24 month period. Following the procedure given in section 2.2.2 we get

$$RR(24 \text{ months}) = 0.13 + 0.92 \times RR(12 \text{ months})$$

with adjusted $R^2 = 0.70$. This model always gives a higher estimated RR after 24 months than after 12 months. This is intuitively correct since we would normally expect RR for an individual not to decrease over time. This model is used to compare the AV&MV model built on the 12 month period with one built on the 24 month period using the procedure described in Section 2.2.2. Results are shown in the third section of Table 1. The 12 month MV model outperforms the 24 month model for

forecasts of 24 month RR for both reported forecast measures. However, this is natural since the recovery period buffer between training and test data (see Figure 2) implies that the 12 month model is built from more recent data. These results indicate that working with a 12 month LGD model is sufficient since the same model can also be used to model longer periods. However, with further investigation, we expect that some hybrid model combining the 12 and 24 month trained LGD models would be the most effective.

5. Conclusion

In the introduction we posed four main research questions. We discuss conclusions relating to each of these questions in turn.

Q1. Which credit card application and default variables are the key drivers of retail LGD?

Our experiments have shown that application variables can be used to model LGD. In particular, Table 2 shows that home status, time with bank and credit bureau score are strong explanatory variables. Additionally, we found that income and balance at default form a joint effect. The negative correlation of default balance with RR matches the findings of Dermine and Neto de Carvalho [2005] that size of loan is a significant explanatory variable for RR. We found that default balance contributes to forecasts of LGD since when it is removed from the AV&MV model, forecast performance is worse. Other variables at time of default such as age of obligor and age of account also influence LGD. Interestingly age has a positive effect on LGD (negative on RR) and we found the linear relation between age and LGD remained

even when age was replaced by several age categories using dummy variables. This is surprising since we would normally expect risk to reduce with maturity. That is, for PD models, typically the positive effect of age peaks in the mid-30's.

Q2. What is the best modelling approach for retail LGD?

Despite trying several different combinations of variables, models such as Tobit and a decision tree and various transformations of the dependent variable – all of which should *in theory* be good models for bimodal LGD – it turns out the best forecast model in our experiments was simple OLS. Why this may be so is unclear, although we conjecture that since LGD is difficult to model, with poor model fit, this implies that regression forecasts tend to fall in a narrow range away from the boundary cases of 0 and 1, therefore models dealing carefully with the boundary cases are superfluous in practice.

Q3. How well do the models perform at forecasting LGD?

Model fit is weak with $R^2 = 0.11$ but such low values are typical of modelling LGD. On a sample of 1118 defaulted financial leases, De Laurentis and Riani [2005] report R^2 values from 0.20 to 0.45 after outliers were deliberately removed. On a data set of 374 defaulted loans to small and medium size firms, Dermine and Neto de Carvalho [2005] report a pseudo R^2 value of 0.13 when considering a 12-month recovery period. The lower R^2 value we report is partly a consequence of the large sample size of our data. In contrast, if we restrict our sample size to just 500 randomly selected cases we get $R^2 = 0.20$. This is more typical of other studies but nevertheless it is misleading and we get adjusted $R^2 = 0.13$ which is closer to the value for the full

sample. Therefore we report adjusted R^2 in our main results and suggest its use for comparison of LGD models across studies with different sample sizes.

For forecasts, the simple model – which effectively forecasts mean LGD from the training data set – does very well and outperforms many of the complex models as can be seen in Table 1. Nevertheless we still see a modest improvement in MSE when model AV&MV compared with the simple model with no variables. For financial institutions, even a small improvement in estimating risk is welcome. When we turn to estimates of LGD across the portfolio however, we can see from Figure 6 that the AV&MV is plainly the better forecaster when compared with the simple model. In this way, these models may prove particularly valuable for estimation of risk at the portfolio level.

Our experiments focussed on modelling LGD for a recovery period of 12 months. However, financial institutions may be interested in longer periods. Nevertheless, we have shown that a 12 month LGD model can be used successfully to forecast 24 month LGD.

Q4. Does the inclusion of MVs lead to improved models of retail LGD?

Our database spanned the period 1999 to 2005. Figure 4 shows that this period covered a range of economic conditions in the UK with interest rates generally decreasing and an overall reduction in unemployment. Earnings generally rose, although at some times growth was higher than others. This period does have the disadvantage for our analysis that there were no major recessions or downturns to train from and towards 2005 the UK economy was stable and fairly unremarkable.

This is a point also noted by Dermine and de Carvalho [2005] with regard to their study. We speculate that a very good macroeconomic model of LGD should have training data across the entire business cycle. Unfortunately, due to practical reasons of data availability, we were unable to provide this. Nevertheless, given this limitation, we still found the MV model to be effective. We show that adding bank interest rates and unemployment level as MVs into a LGD model yields better model fit and that these variables are statistically significant explanatory variables. Additionally including these MVs improves forecasts with generally better MSE and estimates of mean RR across test quarters. Although the improved MSE is modest, Figure 5 suggests that the AV&MV models improve relatively with the duration or size of the training data set. Comparing the AV&MV model with the AV model in Figure 6 shows a clearly better forecast of LGD at the portfolio level. We also report results using the model for separate products where we found that the AV&MV model was less effective, suggesting that several products are required to build effective LGD models based on macroeconomic conditions.

We found that the inclusion of interaction terms between AVs and MVs did not generally improve performance and led to slightly worse results. The poor performance of the model with interaction terms affirms the comment by Gayler [2006] that the main effects are believed to be more stable than interactions for prediction in credit scoring. Nevertheless we feel that there is likely to be some useful interaction effects between MVs and application terms; eg those with high outstanding debt, say a mortgage on a property, are more likely to be affected by changes in bank interest rates. The problem is to determine which of them are important prior to modelling. The automated forward selection process we use is

clearly insufficient for this task. Gayler [2006] recommends that prior expert knowledge is used to determine stable interactions. Therefore useful future work could be conducted to incorporate expert credit advice into the model build stage prior to automated modelling.

Finally, we ran a simple experiment using MV models for stress testing with hypothetical changes in interest rates. For example, we substituted the maximum and minimum interest rate values that occurred during our training period (6% and 3.5%) into the AV&MV model for the last of our test periods, 2005Q2. The forecast mean RR changed by -17% and +24% respectively. These results are plausible in the sense that the forecast mean LGDs were in the range we would expect given historic data. Nevertheless further work is needed to use models with MVs for stress testing. Firstly, a method to calibrate stress test estimates is needed. Secondly the linearity of the models along with the truncated distribution of LGD implies that extreme values of MVs – necessary for stress testing – will have a too-extreme effect on forecasts of LGD. A logit transform of RR would help dampen extreme predictions. However, our experiments show this is not the best model in terms of forecasts. Alternatively, a logit transform of MVs, prior to their use in the model, might also mitigate the problem of extreme forecasts of RR. This is another area of further work.

Acknowledgements

We would like to thank our commercial partners for their assistance and comments in preparing this paper. Research was funded by UK EPSRC grant number EP/D505380/1, working as part of the Quantitative Financial Risk Management Centre.

References

- Akaike, H. (1973). Information theory and an extension of the maximum principle. *Proc. 2nd Int. Symp. Information Theory*, Akademia Kiado, Budapest, 267-281.
- Altman, E.I., Resti, A. and Sironi, A. (2005). Loss given default: a review of the literature. *Recovery Risk* ed. Altman, E., Resti, A. and Sironi, A. (Risk Books).
- Bank for International Settlements BIS (2005). Stress testing at major financial institutions: survey results and practice. *Working report from Committee on the Global Financial System*.
- Basel Committee on Banking Supervision (BCBS 2006). *Basel II: International Convergence of Capital Measurement and Capital Standards*, Basel.
- Bellotti, T. and Crook, J. (2007). Modelling and predicting loss given default for credit cards. *Quantitative Financial Risk Management Centre working paper*.
- Brambor, T., Clark, W.R. and Golder, M. (2005). Understanding interaction models: improving empirical analyses. *Political Analysis* **14**: 63-82.
- De Laurentis, G. and Riani, M. (2005). Estimating LGD in the Leasing Industry: Empirical Evidence from a Multivariate Model. In Altman, E., Resti, A. and Sironi, A. (eds) *Recovery Risk*, Risk Books, London.

Dermine, D. and Neto de Carvalho, C. (2005). Bank loan losses-given-default: a case study. *Journal of Banking and Finance* vol.30, issue 4, 1219-1243

Figlewski, S., Frydman, H. and Liang, W. (2007). Modeling the Effect of Macroeconomic Factors on Corporate Default and Credit Rating Transitions. *NYU Stern Finance working paper*. November 2007.

Financial Services Authority FSA (2005). Stress Testing. *Discussion paper* 05/2 FSA(UK).

Gayler, R. (2006). Comment on “Classifier technology and the illusion of progress – Credit scoring” by Hand, D., *Statistical Science* vol 21, no 1, 19-23.

Grippa, P.S., Iannotti, F. and Leandri, F. (2005) Recovery rates in the banking industry: stylised facts emerging from the Italian experience. In Altman, E., Resti, A. and Sirona, A. (eds) *Recovery Risk*, Risk Books, London.

Gupton, G.M. and Stein, R.M. (2005). *Losscalc V2: dynamic prediction of LGD, modelling methodology*. Moody’s KMV Company.

Hand, D. (2006). Classifier technology and the illusion of progress. *Statistical Science* vol 21, no 1, 1-14.

Kennedy, P. (2003). *A Guide to Econometrics* (5th edition). Blackwell.

Lam, J-P. and Veall, M.R. (2002). Bootstrap prediction intervals for single period regression forecasts. *International Journal of Forecasting* 18 pp 125-130.

Miller, A.J. (1990). *Subset Selection in Regression*. Chapman & Hall NY.

Querci, F. (2005). Loss Given Default on a medium-sized Italian bank's loans: an empirical exercise. *European Financial Management Association* 2005 (Milan, Italy).

Rosche, D. and Scheule, H. (2006) A multifactor approach for systematic default and recovery risk. In Engelmann B and Rauhmeier, R. (eds) *The Basel II Risk Parameters*. Springer, Berlin.

Saurina, J. and Trucharte, C. (2007). *An assessment of Basel II procyclicality in mortgage portfolios*. Banco de España.

Schuermann, T. (2005). What do we know about Loss Given Default? In Altman, E., Resti, A. and Sirona, A. (eds) *Recovery Risk*, Risk Books, London.

Thomas, L.C., Edelman, D.B. and Crook, J.N. (2002). *Credit Scoring and its Applications*. SIAM Monographs on Mathematical Modeling and Computation. SIAM: Philadelphia, USA.

Foot Notes

For reasons of commercial confidentiality, to protect our data supplier, we cannot reveal exact distributions or values of LGD for our data, nor descriptive details of other variables. This is usual in this research area when large portfolios of live financial data are being studied. Nevertheless since our focus is on reporting forecast performance, this should not be a hindrance to scientific reporting.

Tables

Table 1.

Aggregate forecast results for different models.

Experiment					Result	
Recovery period	Product	RR transform	Model	Explanatory variables	MSE	Abs RR diff
12 month	All	None	OLS regression	None (simple)	0.168	0.0384
				AV	0.156	0.0645
				AV & MV	0.151	0.0130
				AV & MV & interaction terms	0.152	0.0169
				AV & MV lag 6	0.155	0.0529
				AV & MV lead 6	0.154	0.0445
			Tobit		0.163	0.0620
			Decision tree		0.221	0.1766
			Least absolute value regression		0.173	0.1144
		Beta		AV & MV	0.166	0.0748
		Logit	OLS regression		0.175	0.1086
		Probit			0.198	0.2041
		Beta	Tobit		0.188	0.1053
			Decision tree		0.192	0.0414
12 month	1			None (simple)	0.160	0.0383
				AV	0.159	0.0795
				AV & MV	0.155	0.0534
	2	None	OLS regression	None (simple)	0.175	0.0391
				AV	0.163	0.0732
				AV & MV	0.195	0.1361
	3			None (simple)	0.165	0.0380
				AV	0.155	0.0774
				AV & MV	0.151	0.0431
	4			None (simple)	0.156	0.0412
				AV	0.149	0.0771
				AV & MV	0.142	0.0227
24 month	All	None	OLS regression		0.194	0.0554
			OLS model built for 12 month recovery period	AV & MV	0.183	0.0525

Abs RR diff is the absolute difference between observed and forecast mean RR over each test quarter. Results are given as mean values across the series of test quarters from 2003Q1 to 2005Q2 or 2004Q2 for 12 or 24 month recovery periods respectively. Figures in bold show best results for each recovery period and product group.

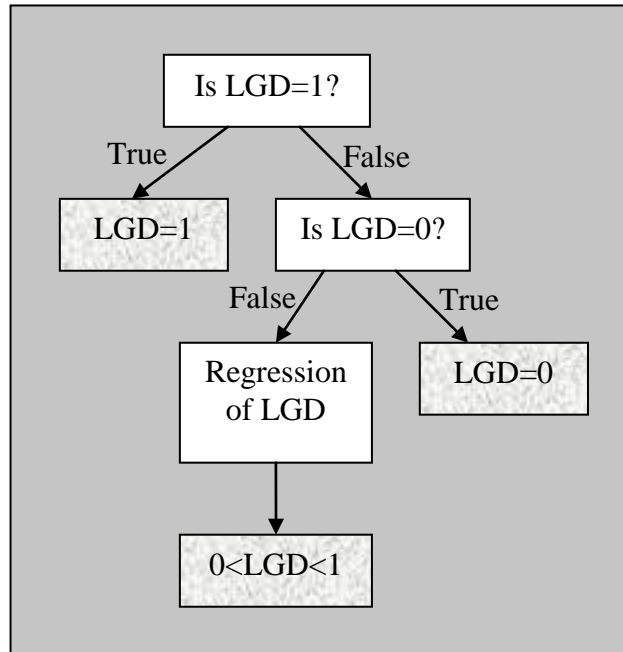
Table 2.
Coefficient estimates for OLS recovery rate model.

Variable	Coefficient estimate	Standard error	z	P> z
Intercept	1.43	0.162	8.83	0
<i>Home status (D):</i> Council tenant	-0.158	0.00544	-29.0	0
Private tenant	-0.146	0.00451	-32.4	0
Term time accommodation	-0.118	0.0410	-2.87	0.004
Other	-0.116	0.00702	-16.6	0
Lives with parents	-0.0643	0.00577	-11.1	0
<i>Excluded category: Own home</i>				
Time with bank	0.0812	0.00716	11.4	0
No data on time with bank (D)	0.0220	0.00774	2.84	0.004
Income (log)	0.0280	0.00770	3.63	0
Income unknown (D)	0.129	0.0328	3.94	0
Number of cards	-0.0153	0.00335	-4.56	0
Years at current address	0.00175	0.000238	7.36	0
<i>Employment status (D):</i> Home maker	-0.0363	0.0109	-3.34	0.001
Unemployed	-0.0338	0.0230	-1.47	0.141
Retired	-0.00570	0.0102	-0.56	0.577
Part time	-0.00451	0.00904	-0.50	0.618
Self employed	-0.00339	0.00525	-0.65	0.519
Other	0.0230	0.00622	3.69	0
Student	0.0266	0.0120	2.21	0.027
<i>Excluded category: Employed</i>				
Age at default	-0.00348	0.000187	-18.6	0
Age unknown (D)	0.0586	0.0146	4.03	0
Credit bureau score	0.186	0.0117	15.9	0
<i>Demographic group (D):</i>				
Council or poor housing area	-0.0404	0.00385	-10.5	0
Rural area	0.0349	0.00933	3.74	0
Suburban or wealthy area	0.0355	0.00498	7.11	0
<i>Excluded category: Other</i>				
<i>Credit card product (D):</i> 1	0.00744	0.00523	1.42	0.155
2	0.00803	0.00660	1.22	0.223
3	0.0147	0.00505	2.90	0.004
<i>Excluded category: 4</i>				
Time on books at default (months)	0.000489	0.000133	3.68	0
Balance at default (log)	-0.165	0.00369	-44.8	0
Date of default	-0.0000161	0.000006	-2.74	0.006
<i>annual effect</i>	-0.00588			
<i>Macroeconomic variables:</i>				
Bank interest rates	-0.0505	0.00296	-17.1	0
<i>standardized estimate</i>	-0.0404			
Unemployment level	-0.000191	0.000048	-3.94	0
<i>standardized estimate</i>	-0.0195			
Earnings growth (log)	0.839	0.484	1.73	0.083
<i>standardized estimate</i>	+0.0032			

Training data set for AV&MV model included all available cases from 1999 to 2005. Standard errors and p-values are computed with bootstrap estimation with 10000 repetitions. Values in *italics* show standardized estimates for MVs and effect over one year for the coefficient on default date. Dummy variable taking the value 0 or 1, for false and true respectively, are indicated by (D).

Figures

Figure 1.
Decision tree model.



The first two binary decision boxes (white) are modelled by logistic regression, whilst the final decision box models LGD using OLS regression.

Figure 2.
Distribution of recovery rates following application of a beta distribution transformation¹.

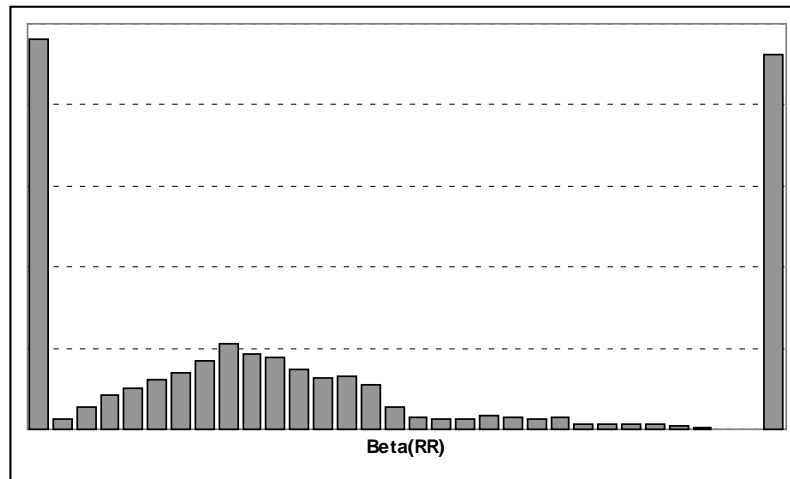
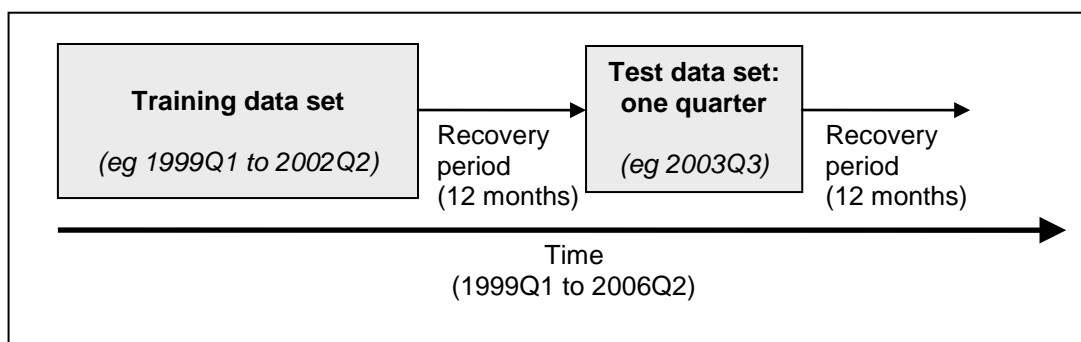


Figure 3.
Illustration of hold out test procedure for one quarter of test data.



The test data set holds all cases that default in the same financial quarter and the training data set comprises all cases that default and have full recovery period prior to the test quarter.

¹ Axes values withheld for reasons of commercial confidentiality of data provider.

Figure 4.

UK macroeconomic data 1999-2005.

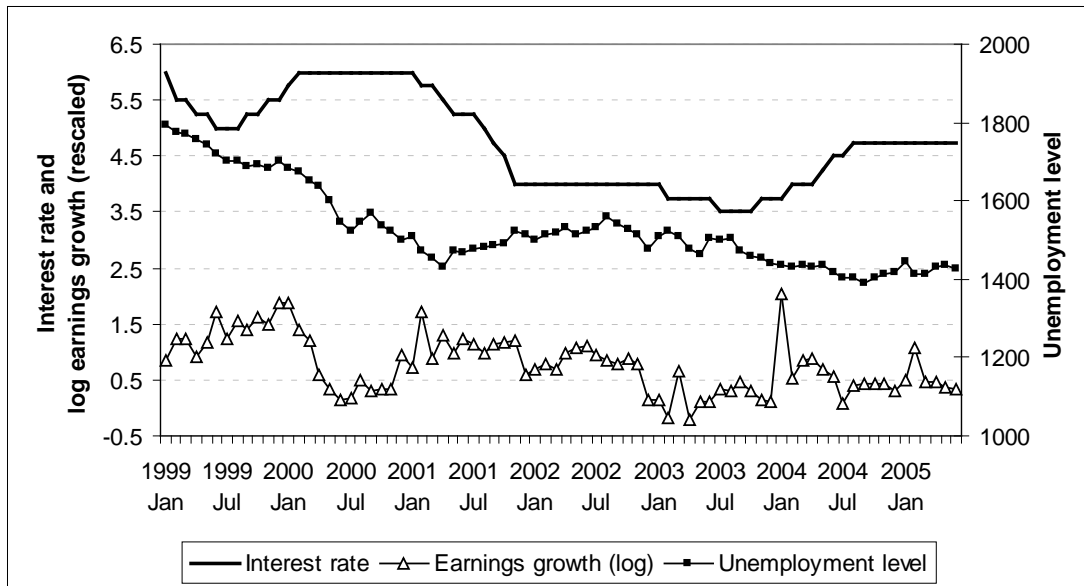


Figure 5.

MSE of forecasts for each test data set from 2003Q1 to 2005Q2 for three OLS models with different variables.

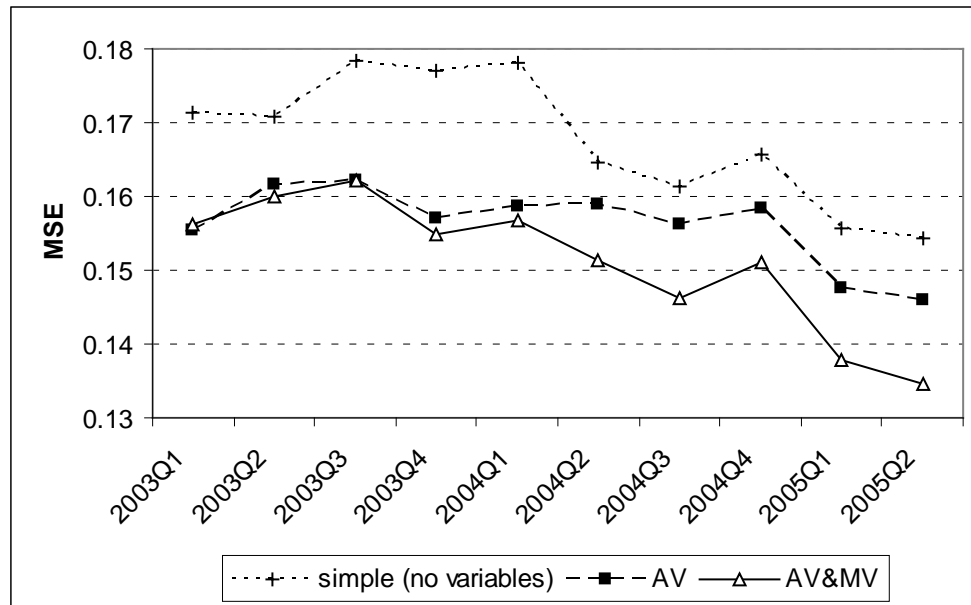
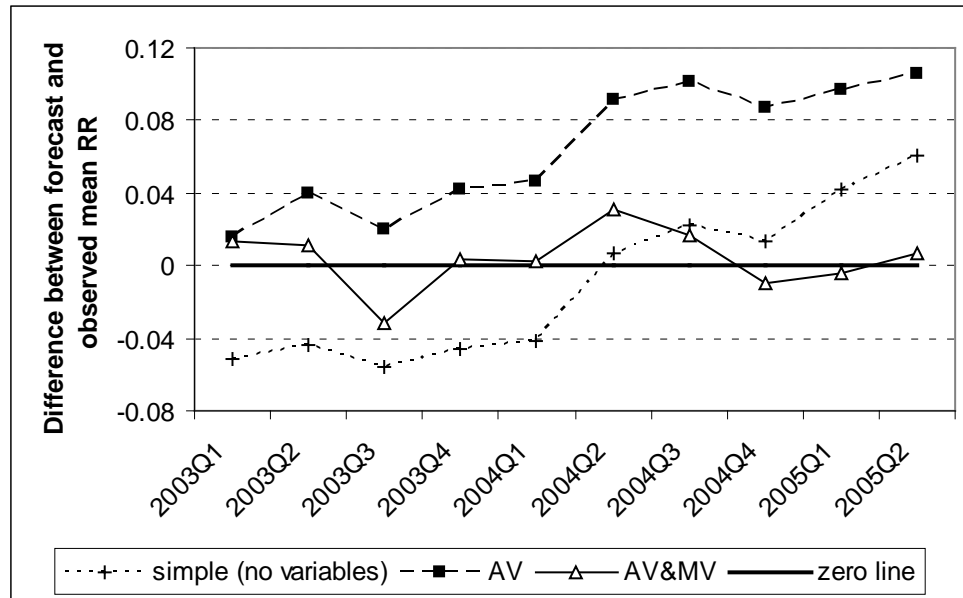


Figure 6.

Difference between forecast and observed mean RR.



Results relate to each test data set from 2003Q1 to 2005Q2 for three OLS model with different variables.